

S1 EP7 - Accelerating the Data Infrastructure Transformation

Thursday, December 16, 2021 · 12:27

John Sakamoto, Vice President, Infrastructure Processors talks OCTEON® 10 with Podcast host Chris Banuelos. In this episode, hear John's insights, industry trends and how Marvell pioneered OCTEON 10. Enabling a multitude of applications including cloud, 5G wireless, enterprise, carrier and datacenter, the OCTEON 10 DPU platform features the industry's first processor family based on a 5nm ARM Neoverse N2 Platform. Tune in to hear John discuss how Marvell is moving the DPU market forward and how Marvell is adapting to market needs as more network workloads become virtualized.

Speaker

John Sakamoto

Vice President, Infrastructure Processors

Host

Christopher Banuelos

Senior Manager of Global Social Media Marketing

C Christopher Banuelos 00:04

Welcome to the Marvell Essential Technology Podcast. I'm your host, Chris Banuelos. On today's episode, I'm with John Sakamoto vice president of infrastructure processors, and today we are discussing all things OCTEON 10. Here John's insights here at industry trends, and see how Marvell is adapting to market needs as more network workloads become virtualized. John, tell me a little bit about your role here at Marvell, and what are some of the kinds of things that you and your team are working on?

J John Sakamoto 00:36

I'm responsible for all the marketing and business development for our processors at Marvell. And we have a really cool role. And we were responsible for educating and evangelizing the value of GPUs for data center, compute and applications. And we get you know, our job is different every day, we get to work with different customers to solve their system challenges, really across a broad spectrum of applications, including enterprise, cloud, 5G, and automotive. And we also kind of look for as well as a team and say, where are things going to be in three to five years and making sure what, what we have in that timeframe is really meeting what our customers need.

C Christopher Banuelos 01:16

John, I want to learn more about your start. And what led you up to Marvell? Can you talk about the early parts of your career?

J John Sakamoto 01:23

Yeah, so I, I back, you know, most of my experience has been in semiconductors and specifically, I spent a long time actually in the field of FPGAs, working for Altera. And it was FPGAs was a really fun industry to be in because, much like the processor business, we got to talk to a lot of different industries, a lot of different

customers across communication, industrial automotive, some of the defense industry. So it was it was a great experience because of the variety of applications that use FPGAs. So I was in the FPGA business for many, many years. Altera was acquired by Intel, and 2015. I worked for Intel for almost four years and was introduced to processors and how FPGAs actually work with processors. And about two years ago, I had the opportunity to move to Marvell I was looking forward to getting back to kind of a little bit smaller company, but you know, wanted to stay in something that was you know, highly programmable, like processors and also kind of a field. That's interesting, because we, you know, processors are used again across a broad spectrum of applications. So, that was one of the main reasons I was attracted to come to Marvell to continue that experience.

C Christopher Banuelos 02:37

So Marvell pioneered the DPU. Can you explain to our listeners what a DPU is? And how it evolves with OCTEON 10?

J John Sakamoto 02:45

Yeah, that's correct. So let you know the definition of DPU is it's a mix of compute accelerators and high speed IO. And really, that is a category of product that was pioneered by Marvell over 20 years ago. And it's something we continue to pioneer. 20 years ago, we recognize that, you know, to increase network system performance, we needed to really combine compute with programmable accelerators to get the performance people needed. And it really started off to kind of process and encrypt data traffic. And you know, the first GPUs we developed were extensively used in firewalls, switches and routers. And since then the Octeon DPU, adoption has really expanded to include cloud, wireless, carrier and enterprise applications. So it really is a broad use case now that that use DPUs and we see that continued to expand.

C Christopher Banuelos 03:41

What have you guys done in OCTEON 10 to move the DPU market forward?

J John Sakamoto 03:47

OCTEON 10 represents a huge move forward for the DPU market, you know, on the compute from the OCTEON 10s built on a TSMC, five nanometer process, and uses the latest and greatest ARM neoverse N2 core. So we're the first to use five and the N2 core and actually have silicon available for that. And really, by combining those two things, this technology represents a 3x jump and compute performance and 50% lower power than previous generations, the DPU. And it's not just only the compute, but we also innovated on the accelerator. So we've added for the first time in a DPU, there's a hard ML [Machine Learning] block that does inline ML processing. We've added hardware based VPP processing, we've added a one terabit switch to the product. And we've done enhancements where we can support true inline crypto. The device also incorporates 56 Gig PAM4 IO, we have PCI Express Gen five support, and we support DDR. So again, when we talked about Compute accelerators High Speed IO is a definition for DPU we've we've made innovations and kind of push the game forward and all three category so that and this is not a device that has, you know, years to come. We're actually now sampling to customers today on this on this product, we have a 24 core version that's gone out to several customers today. So it is here today and will continue to expand that sample distribution to customers as we go through the remainder of this year.

C Christopher Banuelos 05:18

What are some of the industry trends you've observed over the last couple of years? And how is OCTEON 10 leading the way?

J John Sakamoto 05:25

So the need for data center compute continues to be broad? A lot of times people think of DPUs as just being a it's a smartNIC to doing some acceleration in the cloud. And it really is beyond that. Okay, like I mentioned that the need for data center compute is required in the enterprise 5G wireline access even now in automotive, we're starting to see that and this idea of securing network traffic, you know, increasing use of ML. And increased use of things like vector packet processing are areas that our customers expect and, and we focused on innovating there and adding that capability, enhancing that capability in our DPU families.

C Christopher Banuelos 06:08

And how is Marvell adapting to market needs as more network workloads are being virtualized?

J John Sakamoto 06:15

Yeah, in the early days of kind of network virtualization, okay, basically, everyone tried to run these network functions and software only. And that was an easy approach to move over. But what people are finding out very quickly is that to really meet the system performance, and cost requirements, and to keep up with the data rates, that it's just not possible to run, you know, these network functions and software only. So people have kind of quickly realized that, you know, you have to add some accelerators to make it more more effective, more cost effective, you can't just keep adding servers, that doesn't work. So really, to address the challenge, okay, we've been working with our customers to really deliver, you know, DPUs on PCI Express cards, think of them as accelerator cards that will complement software running on a host. And, you know, really what they're using DPUs is to accelerate the the data path functions, okay. So for example, like a firewall, a firewalls job is to kind of look at the flows and determine which flows are safe and which flows need to be denied coming into a network. One of the nice things about using a DPU is that you can process a lot of that flow traffic through the DPU without actually having to go through the hosts. So today, we're we're, you know, recording, you know, a podcast over a zoom call like that traffic flow coming into the network is a safe flow, it's not a threat vector. And so, you know, a DPU Accelerator card can let that flow go through, and you don't have to tax the the actual compute or the host compute to process that workload. And so by combining, you know, DPU accelerator cards with with kind of the host processing, really, we can dramatically lower the cost of that virtualized instance of the firewall. But we can also really increase the performance and make it very competitive to kind of the traditional hardware version of it. So it is an area we're focused on, we see a lot of pickup in customers. It's not just in that enterprise example, you see the same thing and things like 5g and the move to things like vRAN again wireless, you know, running those wireless functions like L1 offload or L2 forwarding, you can't just run that on software, you're going to need your accelerator cards and DPUs on accelerator cards to make those functions run at speed. And to give you that same experience you would get on the traditional RAN network.

C Christopher Banuelos 08:35

And what about the tools to support this development?

J John Sakamoto 08:37

Yeah, that's a very important you know, we have this great silicon will always great capability, you have to make it accessible by the engineers and, and so we have a great tool called Velox, and really, it's designed to help our customers, you know, design with our DPUs. And the one key thing for us is Velox is an open development platform which which leverages the ARM ecosystem, we're not trying to create this closed system and lock people in. And we're really the Velox toolset really has a rich suite of comprehensive networking, security storage stacks, we have comprehensive support for DB DK and VPP. Extensions. And then we have also, you know, very solid support for virtualization supporting things like containers. So again, a very open platform, it's meant to be very, you know, make it very easy for our users to write applications that utilize all the goodness in GPUs. And that's a big part of what we focus on. And a big part of our value proposition is not only combining the great silicon we have, but also that Velox development environment to make it easy to access that silicon.

C Christopher Banuelos 09:44

The last question that I have today is really wanted to focus on what are some of the customer pain points that your team is observing and how is Marvell taking those pain points and turning them into opportunities?

J John Sakamoto 09:57

Yeah, I think there's you know, there's really common pain points across the network that we see from our customers. The first is security that just remains, you know, something that's becoming more and more important, especially as you have more remote workers that the security has to extend, you know, beyond the enterprise into the into the broader network, you know, performance increases are required data rates continue

to rise. And so they have to kind of support those data rates, but they can't necessarily have a higher cost in the solution that they're building. And then there's more and more usage of ML that's required in these network, these network applications. So we're really doing a number of things like security is a major focus for us. You know, in our OCTEON 10 family, which we just announced, you know, we've made major enhancements to kind of our security blocks that really enable things like IPsec and SSL processing to be utilized very easily within the device, but also at higher data rates. So supporting, you know, those data range from 50, up to 400 gig, we mentioned kind of faster data rates, one of the things that we're encouraging our customers to do is look at vector packet processing. The whole concept with that as our hardware within OCTEON. 10 enables the packet processing to be processed as vectors versus doing it packet by packet. And by doing that we can get up to, you know, a 5x performance increase versus our previous generations. And then lastly, again, I mentioned ML, you people are using ML for things like threat detection or Content Aware service delivery. And yes, you can run that in software. But again, it's becoming more and more pervasive. And so one of the reasons that we put a hardware block into OCTEON 10, was to actually accelerate that ML processing. And you can take, you know, algorithms that are running in software and run it into this hardware block and get up to 100x performance. And so again, these are kind of the things we're doing to kind of push the DPU market forward to make sure that as they're more and more challenges from our customers that we're starting to add features in device to make it easier to solve those problems.

C Christopher Banelos 12:01

John, just want to say thank you so much for participating on today's episode, super excited about OCTEON 10. And look forward to doing part two at some point in the future.

J John Sakamoto 12:10

Sounds good, Chris, thank you very much.

C Christopher Banelos 12:14

Thank you for listening to the Marvell Essential Technology Podcast. As always, please feel free to visit our website to learn more, and we'll see you on the next episode.



To deliver the data infrastructure technology that connects the world, we're building solutions on the most powerful foundation: our partnerships with our customers. Trusted by the world's leading technology companies for 25 years, we move, store, process and secure the world's data with semiconductor solutions designed for our customers' current needs and future ambitions. Through a process of deep collaboration and transparency, we're ultimately changing the way tomorrow's enterprise, cloud, automotive, and carrier architectures transform—for the better.

Copyright © 2022 Marvell. All rights reserved. Marvell and the Marvell logo are trademarks of Marvell or its affiliates. Please visit www.marvell.com for a complete list of Marvell trademarks. Other names and brands may be claimed as the property of others.